

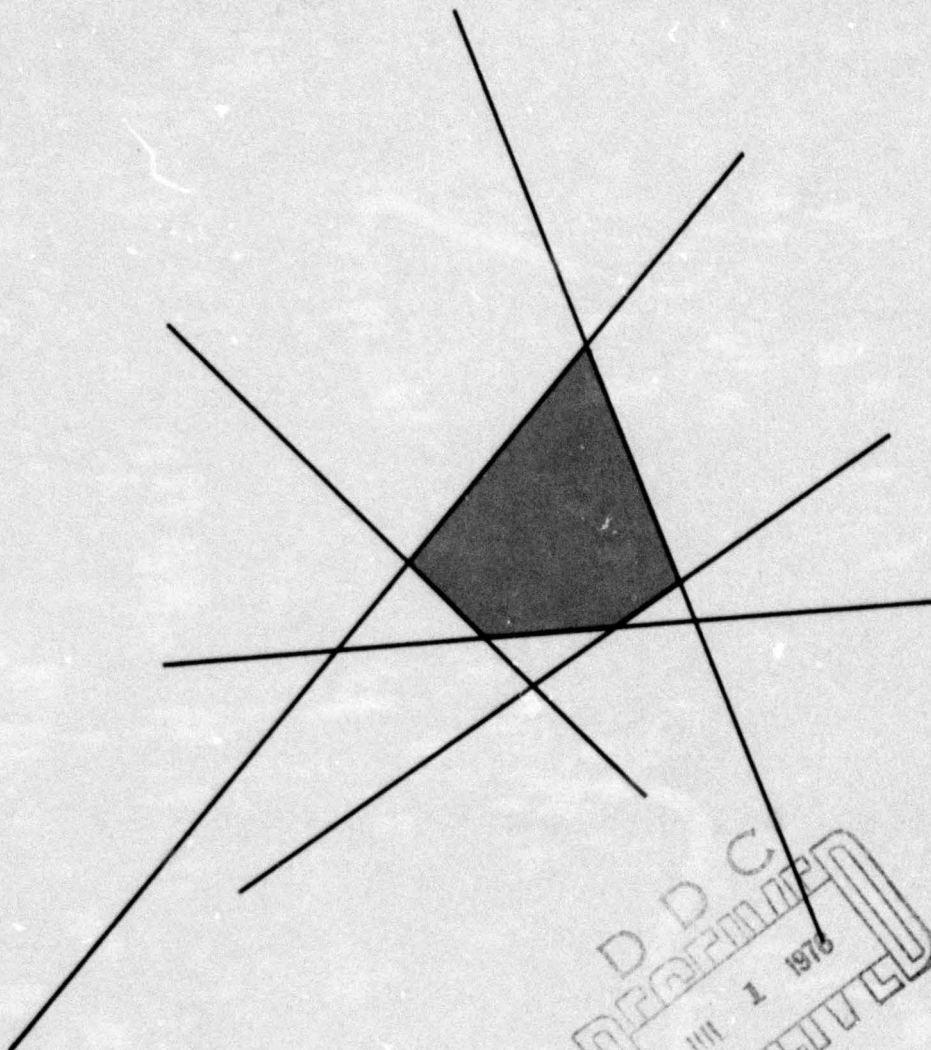
FC

# APPROXIMATIONS IN MULTI-SERVER POISSON QUEUES

12

by  
SHIRLEY A. NOZAKI  
and  
SHELDON M. ROSS

ADA 026304



OPERATIONS  
RESEARCH  
CENTER

DDC  
APPROVED  
MAY 1 1976  
REGISTERED

DISTRIBUTION STATEMENT A  
Approved for public release;  
Distribution Unlimited

UNIVERSITY OF CALIFORNIA • BERKELEY

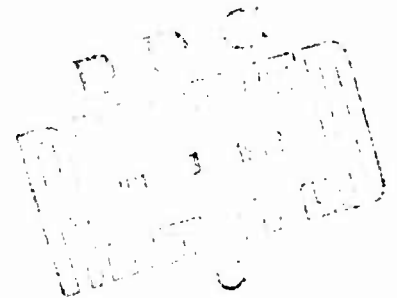
APPROXIMATIONS IN MULTI-SERVER POISSON QUEUES

by

Shirley A. Nozaki  
Operations Research Center  
University of California, Berkeley

and

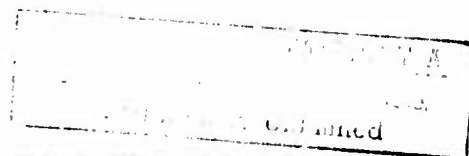
Sheldon M. Ross  
Department of Industrial Engineering  
and Operations Research  
University of California, Berkeley



APRIL 1976

ORC 76-10

This research has been partially supported by the Office of Naval Research under Contract N00014-75-C-0781 and the U. S. Army Research Office - Research Triangle Park under Grants DAHCO4-75-G-0163 and DAAG29-76-G-0042 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.



Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC-76-10	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) APPROXIMATIONS IN MULTI-SERVER POISSON QUEUES.		5. TYPE OF REPORT & PERIOD COVERED Research Report.
7. AUTHOR(s) Shirley A./Nozaki Sheldon M./Ross		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-238
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, Virginia 22217		12. REPORT DATE Apr 1976
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 23
(12) 24p.		15. SECURITY CLASS. (of this report) Unclassified
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited.		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES  Also supported by the U. S. Army Research Office - Research Triangle Park under Grants DAHC04-75-G-0163 and DAAG29-76-G-0042.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Poisson Arrivals General Service Distribution Average Wait in Queue Approximation Assumption Mini Service		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  (SEE ABSTRACT)		

# ABSTRACT

→ Our major objective is to obtain an approximation for the average time spent waiting in queue by a customer in an M/G/k queueing system--call it  $W_Q$ . This is done by means of an approximation assumption presented, ~~in Section 2~~, which is shown to be asymptotically valid both in heavy and in light traffic. ~~In Section 3~~, the approximation assumption is used to derive an approximation for  $W_Q$ . Numerical comparison with tables given by Hillier-Lo in the special case of Erlang service times indicate that the approximation, which depends on the service distribution only through its first two moments, works remarkably well. In addition, as a by-product of our analysis, we also obtain approximations for the distribution of the number of busy servers and the mean length and number of customers in a busy period. These latter approximations depend on the service distribution only through its mean.

In Section 4, ~~we show that~~ the approximation assumption is valid and leads to the exact result in the case of a limited capacity system where no queue is allowed to form. ✓

## APPROXIMATIONS IN MULTI-SERVER POISSON QUEUES

by

Shirley A. Nozaki and Sheldon M. Ross

### 0. INTRODUCTION AND SUMMARY

In this paper we consider an M/G/k queueing system - that is a system in which customers arrive in accordance with a Poisson process having rate  $\lambda$ , and are serviced by one of  $k$  servers, each of whom has service distribution  $G$ . Upon arrival a customer will either enter service if at least one server is free or else join the queue if all servers are busy. Our results will be independent of the order of service of those waiting in queue as long as it is supposed that a server will never remain idle if customers are waiting. To facilitate the analysis, however, we will suppose a service discipline of "first come first to enter service."

Our major objective is to obtain an approximation for the average time spent waiting in queue by a customer - call it  $W_Q$ . This is done by means of an approximation assumption presented in Section 2, which is shown to be asymptotically valid both in heavy and in light traffic. In Section 3 the approximation assumption is used to derive an approximation for  $W_Q$ . Numerical comparison with tables given by Hillier-Lo in the special case of Erlang service times indicate that the approximation, which depends on the service distribution only through its first two moments, works remarkably well. In addition, as a by-product of our analysis, we also obtain approximations for the distribution of the number of busy servers and the mean length and number of customers in a busy period. These latter approximations depend on the service distribution only through its mean.

In Section 4 we show that the approximation assumption is valid and leads to the exact result in the case of a limited capacity system where no queue is allowed to form.

Future research plans are indicated in Section 5.

Throughout this paper we suppose that

$$\lambda \int_0^{\infty} x dG(x) < k$$

and

$$\int_0^{\infty} x^2 dG(x) < \infty .$$

## 1. BASIC DEFINITIONS AND FUNDAMENTAL EQUATION

We shall need the following notation:

- $P_1$ : the steady state probability that there are 1 people in the system
- $S$ : a service time random variable, i.e.,  $P\{S \leq x\} = G(x)$
- $W_Q$ : the average amount of time that a customer spends waiting in queue (does not include service time)
- $L_Q$ : the (time) average number of customers waiting in queue
- $V$ : the (time) average amount of work in the system, where the work in the system at any time is defined to be the total (of all servers) amount of service time necessary to empty the system of all those presently either being served or waiting in queue.

We will make use of the following idea (previously exploited in such papers as [1], [2] and [5]) that if a (possibly fictitious) cost structure is imposed, so that customers are forced to pay money (according to some rule) to the system, then the following identity holds - namely

$$(1) \quad \begin{aligned} & \text{(time) average rate at which the system earns money} \\ &= \text{average arrival rate of customers} \times \text{average amount} \\ & \quad \text{of money paid by a customer.} \end{aligned}$$

A heuristic proof of the above is that both sides of (1) times  $T$  is approximately equal to the total amount of money paid to the system by time  $T$ , and the result follows by dividing by  $T$  and then letting  $T \rightarrow \infty$ .<sup>†</sup>

<sup>†</sup> A rigorous proof along these lines can easily be established in the models we consider since all have regeneration points. More general conditions under which it is true are presented in [1].

By choosing appropriate cost rules many useful formulas can be obtained as special cases of (1). For instance by supposing that each customer pays \$1 per unit time while in service, Equation (1) yields that

$$(2) \quad \text{average number in service} = \lambda E[S] .$$

Similarly by supposing that each customer pays \$1 per unit time while waiting in queue, we obtain from (1) that

$$(3) \quad L_Q = \lambda W_Q .$$

Also, if we suppose that each customer in the system pays \$x per unit time whenever its remaining service times is x, then (1) yields that

$$\begin{aligned} (4) \quad V &= \lambda E \left[ SW_Q^* + \int_0^S (S - x) dx \right] \\ &= \lambda \left[ E(SW_Q^*) + E(S^2)/2 \right] \\ &= \lambda E[S]W_Q + \lambda E[S^2]/2 \end{aligned}$$

where  $W_Q^*$  is a random variable representing the (limiting) amount of time that the n-th customer spends waiting in queue.

Equation (4) will be of particular use to us.

Another important fact which we shall use is that, since our arrival stream of customers is a Poisson process, the probability structure of what an arrival observes is identical to the steady state probability structure of the system.<sup>†</sup> Thus, for instance  $\lambda E[S]$  will equal the average number of busy servers that an arrival observes;  $V$ , will equal the average amount of

<sup>†</sup> Intuitively this is so since for a Poisson arrival process (a) the distribution of the times at which arrivals occur is uniform, and (b) given that an arrival occurs at time t, the conditional distribution for the remaining arrivals is the same distribution as for the original Poisson process.

work in the system as seen by an arrival; and  $P_i$ , the probability that an arrival finds  $i$  people presently in the system.

As a result of the above we may write Equation (1) as

$$\begin{aligned}
 & \text{Average rate at which system earns money} \\
 (5) \quad & = \lambda \sum_i P_i \times \text{average amount paid by a customer finding} \\
 & \quad i \text{ people already in the system when he arrives.}
 \end{aligned}$$

## 2. THE APPROXIMATION ASSUMPTION

Let  $G_e$  denote the equilibrium distribution of  $G$ . That is,

$$G_e(x) = \int_0^x \frac{(1 - G(y))}{E[S]} dy ,$$

also let

$$\delta(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} .$$

We shall make the following approximation assumption.

### Approximation Assumption:

Given that a customer arrives to find  $i$  busy servers,  $i > 0$ , then at the time that he enters service, the remaining service times of the other  $i - \delta(i,k)$  customers being served has a joint distribution that is approximately that of independent random variables each having distribution  $G_e$ .

### Heuristic Remarks Concerning the A.A.:

1. First we note that the A.A. is asymptotically true either in heavy traffic (that is, as  $\lambda E[S] \rightarrow k$ ) or in light traffic (that is, as  $\lambda E[S] \rightarrow 0$ ). This is so in heavy traffic since the great majority of arrivals will encounter a large queue and as a result the  $k$  departure processes (one for each server) they observe will be approximately independent delayed renewal processes. Hence, considering those customers served by server  $i$ , it follows that when they enter service they would have been observing  $k - 1$  independent delayed renewal processes for a large time, and the A.A. follows since the limiting distribution of excess in a renewal process is just  $G_e$ .

In extremely light traffic the great majority of arrivals will find either 0 or 1 busy servers. Now since Poisson arrivals see the system as it is (averaged over all time) it follows that arrivals finding 1 server busy would encounter the same additional service time (for the busy server) as would random (and uniform) time sampling of the excess of a renewal process. Hence, the A.A. follows in light traffic from the renewal process (excess) result.

2. In fact the same reasoning given above for light traffic shows that if we isolate attention upon a particular server, then whenever customers arrive to find this server busy, the remaining service time will have distribution  $G_e$ . Now it should be noted that this is not the same as saying (as the A.A. does) that the remaining service time of this server will have distribution  $G_e$  at the moment when a customer is about to *enter* service (as opposed to when he *arrives*). However we might hope that it should be close.
3. Additional heuristics for the A.A. follows from the fact that it is known to be (exactly) true when no queue is allowed (see Section 4).

### 3. THE APPROXIMATION

For any arbitrary arrival we have the following identity:

$$\begin{aligned} & \text{work in the system at the time of his arrival} \\ &= k \times \text{time he spends waiting in queue} + R \end{aligned}$$

where

$R$  = sum of the remaining service times of all those being served at the time when the arrival enters service.

Taking expectations yields, since a Poisson arrival sees the system as it is in steady state, that

$$V = kW_Q + E[R] .$$

To obtain  $E[R]$  we condition on  $B$ , the number of servers that are busy when the customer arrives.

$$E[R] = E[E(R/B)]$$

$$= E[B - \delta(B,k)] \frac{E[S^2]}{2E[S]}$$

where the last equation follows from the A.A., and the fact that  $\int_0^\infty x dG_e(x) = E[S^2]/2E[S]$ . Since  $E[B] = \lambda E[S]$  from Equation (2), we obtain that

$$\begin{aligned} E[R] &= \lambda \frac{E[S^2]}{2} - \frac{E[S^2]}{2E[S]} P\{B = k\} \\ &= \lambda \frac{E[S^2]}{2} - \frac{E[S^2]}{2E[S]} \bar{P}_k \end{aligned}$$

where

$$\bar{P}_k = 1 - \sum_{i=0}^{k-1} P_i .$$

Hence,

$$V = kW_Q + \lambda \frac{E[S^2]}{2} - \frac{E[S^2]}{2E[S]} \bar{P}_k.$$

However, from Equation (4) we know that

$$V = \lambda E[S]W_Q + \lambda \frac{E[S^2]}{2}$$

implying that

$$(6) \quad W_Q = \frac{E[S^2] \bar{P}_k}{2E[S](k - \lambda E[S])}.$$

Thus we need  $\bar{P}_k$ . To obtain the probability distribution of the number of busy servers we impose the following fictitious cost structure - namely that the  $i$  oldest customers in the system pay \$1 per unit time,  $i = 1, 2, \dots, k$ , where the age of a customer is measured from the moment it enters the system. Hence, letting  $S_1^e, S_2^e, \dots, S_{k-1}^e$  denote  $k-1$  independent random variables each having distribution  $G_e$  we obtain from Equation (5) that

$$\begin{aligned} & P_1 + 2P_2 + \dots + (i-1)P_{i-1} + i(1 - P_0 - \dots - P_{i-1}) \\ &= \lambda(P_0 + \dots + P_{i-1})E[S] + \lambda P_i E\left[\left(S - \min(S_1^e, S_2^e, \dots, S_i^e)\right)^+\right] \\ &+ \lambda P_{i+1} E\left[\left(S - 2\text{nd smallest of } (S_1^e, \dots, S_{i+1}^e)\right)^+\right] \\ &+ \vdots \\ (7) \quad &+ \lambda P_{k-2} E\left[\left(S - (k-1-i)\text{th smallest of } (S_1^e, \dots, S_{k-2}^e)\right)^+\right] \\ &+ \lambda(1 - P_0 - \dots - P_{k-2}) E\left[\left(S - (k-i)\text{th smallest of } (S_1^e, \dots, S_{k-1}^e)\right)^+\right] \\ & \qquad \qquad \qquad i = 1, \dots, k-1 \\ & P_1 + 2P_2 + \dots + (k-1)P_{k-1} + k(1 - P_0 - \dots - P_{k-1}) = \lambda E[S] \end{aligned}$$

$\left( \text{where } x^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \right)$ . To understand the above equations suppose first that  $i < k$ . Now as only the  $i$  oldest pay it follows that when  $j$  customers are present the system earns at a rate  $j$  when  $j \leq i$  and at a rate  $i$  when  $j > i$ . Hence the left side of Equation (7) represents the average rate at which the system earns. On the other hand an arrival finding fewer than  $i$  customers already in the system will immediately go into service and will pay a total amount equal to his service time; while an arrival finding  $j$  present,  $k - 1 \geq j \geq i$  will also go immediately into service but will only begin paying when  $j - i + 1$  of the  $j$  others in service leave. Thus in this latter case, under the A.A., the arrival would expect to pay a total of  $E \left[ \left( S - (j + 1 - i) \text{th smallest of } (S_1^e, S_2^e, \dots, S_j^e) \right)^+ \right]$ . Finally if the arrival found more than  $k - 2$  busy servers then he will begin paying after  $k - i$  of those customers in service when he enters service leave the system. This explains the first  $k - 1$  of the set of Equation (7). The last equation (when  $i = k$ ) easily follows since in this case each customer will pay a total equal to his time in service.

To simplify the set of Equation (7) we will need the following lemma.

Lemma 1:

If  $S, S_1^e, \dots, S_r^e$  are independent random variables such that  $S$  has distribution  $G$  and the others  $G_e$ , then

$$E \left[ \left( S - j \text{th smallest of } (S_1^e, \dots, S_r^e) \right)^+ \right] = \frac{r + 1 - j}{r + 1} E[S].$$

Proof:

Using the identity

$$(x - y)^+ = x - \min(x, y)$$

we have that

$$\begin{aligned} & E\left[\left(S - j\text{th smallest of } S_1^e, \dots, S_r^e\right)^+\right] \\ &= E[S] - E\left[\min\left(S, j\text{th smallest of } S_1^e, \dots, S_r^e\right)\right]. \end{aligned}$$

Now,

$$\begin{aligned} & E\left[\min\left(S, j\text{th smallest of } S_1^e, \dots, S_r^e\right)\right] \\ &= \int_0^\infty P\{S > a\} P\left\{j\text{th smallest of } (S_1^e, \dots, S_r^e) > a\right\} da \\ &= \int_0^\infty (1 - G(a)) \sum_{i=0}^{j-1} \binom{r}{i} (G_e(a))^i (1 - G_e(a))^{r-i} da \\ &= E[S] \sum_{i=0}^{j-1} \binom{r}{i} \int_0^1 y^i (1-y)^{r-i} dy \quad \left(\begin{array}{l} \text{by the substitution} \\ y = G_e(a) \\ dy = \frac{(1 - G(a))}{E[S]} da \end{array}\right) \\ &= E[S] \sum_{i=0}^{j-1} \binom{r}{i} \frac{i!(r-i)!}{(r+1)!} \\ &= E[S] \frac{j}{r+1} \end{aligned}$$

which proves the lemma. ■

Hence, using Lemma 1, the Equation (7) become

$$\begin{aligned} & P_1 + 2P_2 + \dots + (i-1)P_{i-1} + i(1 - P_0 - \dots - P_{i-1}) \\ &= \lambda(P_0 + \dots + P_{i-1})E[S] + \lambda P_i \frac{1}{i+1} E[S] + \dots + \lambda P_{k-2} \frac{1}{k-1} E[S] \\ (8) \quad & + \lambda(1 - P_0 - \dots - P_{k-2}) \frac{1}{k} E[S], \quad i = 1, \dots, k-1 \\ & P_1 + 2P_2 + \dots + (k-1)P_{k-1} + k(1 - P_0 - \dots - P_{k-1}) \\ &= \lambda E[S]. \end{aligned}$$

Now, as the above equations for the  $P_i$ ,  $i = 0, 1, \dots, k-1$ , depend only on  $G$  through  $E[S]$  and as the equations are exactly true when  $G$  is exponential (since the A.A. is exact when  $G$  is exponential) it follows (since it can be shown that the set of equations has at most one solution) that the solution of (8) is identical to the well known solution in the case M/M/k. That is,

$$P_0 = \left[ \sum_{n=0}^{k-1} \frac{(\lambda E[S])^n}{n!} + \frac{(\lambda E[S])^k}{(k-1)!(k - \lambda E[S])} \right]^{-1}$$

and

$$P_i = \frac{(\lambda E[S])^i}{i!} P_0, \quad i = 1, \dots, k-1.$$

Hence,

$$\bar{P}_k = \frac{(\lambda E[S])^k}{(k-1)!(k - \lambda E[S]) \left[ \sum_{n=0}^{k-1} \frac{(\lambda E[S])^n}{n!} + \frac{(\lambda E[S])^k}{(k-1)!(k - \lambda E[S])} \right]}$$

and our approximation for  $W_Q$  is thus given by

$$(9) \quad W_Q = \frac{\lambda^k E[S^2] (E[S])^{k-1}}{2(k-1)!(k - \lambda E[S])^2 \left[ \sum_{n=0}^{k-1} \frac{(\lambda E[S])^n}{n!} + \frac{(\lambda E[S])^k}{(k-1)!(k - \lambda E[S])} \right]}.$$

In the special case of two servers,  $k = 2$ , we have that

$$\bar{P}_2 = \frac{(\lambda E[S])^2}{2 + \lambda E[S]} \quad \text{when } k = 2$$

and (9) reduces to

$$W_Q = \frac{\lambda^2 E[S^2] E[S]}{2(2 - \lambda E[S])(2 + \lambda E[S])} \quad \text{when } k = 2.$$

Numerical tables for  $L_Q$  have been published by Hillier and Lo in the special case  $M/E_r/k$ , where  $E_r$  represents an Erlang distribution with  $r$  phases. That is, a service time has the same distribution as the sum of  $r$  independent and identically distributed exponential random variables. The following tables compares our approximate formula for  $L_Q$  (namely  $\lambda W_Q$ ) with the Hillier-Lo tables.

It is also interesting to see how our approximation for the probability that all servers are busy (call it  $P\{\text{all busy}\}$ ) compares with the actual values in the special case  $M/E_r/k$ . Again referring to the Hillier-Lo tables we have Table 3.

Remarks:

1. It is interesting to note that in all cases the approximation for  $W_Q$  is slightly less than the exact value in the case of Erlang service times. While the reason for this is by no means apparent and further study is clearly indicated the authors feel that it may be relevant that the Erlang is an increasing failure rate distribution.
2. The approximation for  $P_0$  also leads to approximations for  $E[B]$  and  $E[C]$ , the expectations of the length of and the number of customers served in a busy period. This follows since from the theory of alternating renewal processes we have that

$$P_0 = \frac{1/\lambda}{1/\lambda + E[B]}$$

or

$$E[B] = \frac{1 - P_0}{\lambda P_0}.$$

Also, letting  $X_i$  denote the time between the  $i$ -th and  $(i + 1)$ st arrival then

$$E \left[ \sum_{i=1}^C X_i \right] = E[B] + 1/\lambda .$$

However, by Wald's equation

$$E \left[ \sum_{i=1}^C X_i \right] = E[C]/\lambda$$

and so

$$\begin{aligned} E[C] &= \lambda E[B] + 1 \\ &= 1/P_0 . \end{aligned}$$

TABLE 1

$$L_Q \text{ for } M/E_r/2, L_Q = \frac{r+1}{r} \frac{\rho^3}{1-\rho^2}, \rho = \frac{\lambda E[S]}{2}$$

$\frac{r}{\rho}$	2	3	4	5	6	7	8
.1	.00152 .00161	.00135 .00148	.00126 .00142	.00121 .00138	.00118 .00136	.00115 .00134	.00114 .00133
.3	.04451 .04625	.03956 .04192	.03709 .03976	.03560 .03847	.03462 .03760	.03391 .036986	.03338 .03652
.5	.25 .25564	.22223 .22975	.20834 .21678	.2 .20899	.19445 .20328	.19048 .20005	.18750 .19724
.7	1.0082 1.0201	.89673 .91165	.84061 .85736	.80706 .82474	.78464 .80295	.76863 .78237	.75662 .77566
.9	5.75526 5.7732	5.11579 5.13953	4.79605 4.82255	4.60421 4.63227	4.4765 4.5053	4.38496 4.4146	4.31645 4.3466
.95	13.19038 13.21011	11.72479 11.75088	10.99194 11.02111	10.55231 10.58313	10.25919 10.29108	10.04867 10.08242	9.89279 9.92589
.99	73.13812 73.1593	65.01165 65.0397	60.94843 60.97969	58.51049 58.54358	56.88520 56.91942	55.72427 55.75926	54.85358 54.8891

top number in box = approximation for  $L_Q$ 

bottom number in box = exact value as given by Hillier-Lo

TABLE 2

$$L_Q \text{ for } M/E_r/k, L_Q = \frac{(\rho k)^{k+1} \frac{r+1}{r}}{2(k!)k(1-\rho)^2 \left[ \sum_{n=0}^{k-1} \frac{(\rho k)^n}{n!} + \frac{(\rho k)^k}{k!(1-\rho)} \right]} \quad \rho = \frac{\lambda E[S]}{k}$$

$\rho \backslash k$	3	4	5	6	7	8	9	10
	$\frac{r=2}{r}$	$\frac{r=3}{r}$	$\frac{r=4}{r}$	$\frac{r=2}{r}$	$\frac{r=3}{r}$	$\frac{r=2}{r}$	$\frac{r=2}{r}$	$\frac{r=2}{r}$
.1	.000309	.000274	.000257	.000066	.000015	.000003	.000001	0
	.000341	.000319	.000310	.000075	.000071	.000003	.000001	0
.3	.022509	.020008	.018758	.011909	.006473	.003583	.002009	.000648
	.023993	.022024	.0210483	.0129386	.007142	.004004	.002269	.000745
.5	.177632	.157895	.148026	.130435	.115942	.074357	.057149	.034537
	.184420	.16697	.158227	.1370512	.124807	.079616	.0616708	.048126
.7	.861603	.765869	.718002	.750145	.666796	.587961	.526293	.427911
	.877853	.787394	.7420435	.768936	.6916764	.608581	.547007	.494070
.9	5.515162	4.902366	4.595968	5.317335	4.726520	5.146829	4.859712	4.735348
	5.54407	4.9405	4.638435	5.354183	4.775027	5.189889	5.043990	4.912132
.95	12.924871	11.488775	10.770726	12.702721	11.291307	12.508636	12.334640	12.029403
	12.95731	11.53154	10.81833	12.74479	11.34665	12.55858	12.39130	12.09718
.99	72.851736	64.757099	60.709780	72.609459	64.541742	72.395734	72.202451	71.859417
	72.88707	64.80367	60.76161	72.65587	64.60278	72.45146	72.26631	71.93724

top number in box = approximation for  $L_Q$ 

bottom number in box = exact value as given by Hillier-Lo

TABLE 3

$$p\{\text{delay}\} = \frac{(\rho k)^k}{k!(1-\rho) \left[ \sum_{n=0}^{k-1} (\rho k)^n/n! + \frac{(\rho k)^k}{k!(1-\rho)} \right]} \quad \rho = \frac{\lambda E[S]}{k}$$

$\rho \backslash k$	3	4	5	6	7	8	9	10
.1	.003704	.000794	.000176	.000040	.000009	.000002	0	0
	.003680	.000789	.000174	.000039	.000009	.000002	0	0
.3	.070029	.037050	.020139	.011146	.006249	.003537	.002017	.001157
	.069132	.036479	.019799	.0109501	.006137	.003474	.001981	.0011368
.5	.236842	.173913	.130371	.099143	.076198	.059044	.046050	.036105
	.233828	.17105	.127865	.0970317	.074455	.0576235	.04490	.0351801
.7	.492344	.428654	.377838	.335978	.300739	.270603	.244520	.221731
	.487999	.423547	.372368	.3303820	.295166	.265145	.23924	.216656
.9	.817061	.787753	.762493	.740126	.719957	.701533	.684535	.668732
	.814509	.784352	.758402	.735459	.714804	.695960	.678598	.662475
.95	.907009	.891415	.877799	.865589	.854452	.844169	.834584	.825586
	.905583	.8894738	.874510	.862811	.851329	.8407339	.830865	.8216067
.99	.981168	.977905	.975027	.972424	.970030	.967804	.965714	.963738
	.980858	.977476	.974491	.971793	.9693117	.967004	.964838	.962792

top number in box = approximation for  $P\{\text{delay}\}$ bottom number in box = exact value as given by Hillier-Lo for  $M/E_2/k$

#### 4. NO QUEUE ALLOWED - ERLANG'S LOSS FORMULA

In this section we suppose that arriving customers that find all servers busy are lost to the system. In this case it is known (see [6]) that the A.A. is exact. The limiting probabilities can be obtained by again supposing that the  $i$  oldest customers pay \$1 per unit time while in the system. From the fundamental Equation (5) and Lemma 1 we obtain the equation

$$\begin{aligned}
 & P_1 + 2P_2 + \dots + (i-1)P_{i-1} + i(1 - P_0 - \dots - P_{i-1}) \\
 &= \lambda(P_0 + \dots + P_{i-1})E[S] + \lambda P_1 \frac{1}{i+1} E[S] + \dots + \lambda P_{k-2} \frac{1}{k-1} E[S] + \lambda P_{i-1} \frac{1}{k} E[S], \\
 & \qquad \qquad \qquad i = 1, \dots, k-1
 \end{aligned}$$

$$P_1 + 2P_2 + \dots + kP_k = \lambda(1 - P_k)E[S].$$

The above equations, along with the equation,

$$\sum_{i=0}^k P_i = 1$$

can now be solved to yield the well known result known as Erlang's loss formula - namely

$$P_i = \frac{(\lambda E[S])^i / i!}{\sum_{n=0}^k (\lambda E[S])^n / n!} \qquad i = 0, 1, \dots, k.$$

## 5. FUTURE RESEARCH

In future work the authors are planning to employ the approach of the present paper to obtain an approximation for  $W_Q$  in such extensions as

- (i) finite capacity models
- (ii) batch arrival models
- (iii) models in which each server has a different service distribution.

It is felt that such models have great applicability in the real world.

## REFERENCES

- [1] Brumelle, S. L., "On the Relation Between Customer and Time Averages in Queues," Journal of Applied Probability, Vol. 8, pp. 508-520, (1971).
- [2] Brumelle, S. L., "A Generalization of  $L = \lambda W$  to Moments of Queue Length and Waiting Times," Operations Research, Vol. 20, No. 6, pp. 1127-1136, December, (1972).
- [3] Hillier, F. S. and F. D. Lo, "Tables for Multiple-Server Queueing Systems Involving Erlang Distributions," Technical Report No. 31, Department of Operations Research, Stanford University, December, (1971).
- [4] Kingman, J. F. C., "Inequalities in the Theory of Queues," Journal of the Royal Statistical Society, B, Vol. 32, pp. 102-110, (1970).
- [5] Stidham, S., "Static Decision Models for Queueing Systems with Non-Linear Waiting Costs," Technical Report No. 9, Stanford University, (1968).
- [6] Takács, L., "A Simple Proof of Erlang's Loss Formula," Annals of Mathematical Statistics.